Published in final edited form as:

Gene. 2015 February 10; 556(2): 149–152. doi:10.1016/j.gene.2014.11.049.

Genetic variants and risk of esophageal squamous cell carcinoma: A GWAS-based pathway analysis

Xi Yang^{a,1}, Hongcheng Zhu^{a,1}, Qin Qin^{a,1}, Yuehua Yang^a, Yan Yang^a, Hongyan Cheng^b, and Xinchun Sun^{a,*}

^aDepartment of Radiation Oncology, The First Affiliated Hospital of Nanjing Medical University, Nanjing 210019, China

^bDepartment of General Internal Medicine, The First Affiliated Hospital of Nanjing Medical University, Nanjing 210019, China

Abstract

This study was designed to identify candidate single-nucleotide polymorphisms (SNPs) that may affect the susceptibility to esophageal squamous cell carcinoma (ESCC) and elucidate their potential mechanisms to generate SNP-to-gene-to-pathway hypotheses. A genome-wide association study (GWAS) dataset for ESCC, which included 453,852 SNPs from 1898 ESCC patients and 2100 control subjects of Chinese population, was reviewed. The identify candidate causal SNPs and pathways (ICSNPPathway) analysis identified seven candidate SNPs, five genes, and seven pathways, which together revealed seven hypothetical biological mechanisms. The three strongest hypothetical biological mechanisms were as follows: rs4135113 → TDG → BASE EXCISION REPAIR; rs1800450 → MBL2 → MONOSACCHARIDE BINDING; and rs3769823 → CASP8 → d4gdiPathway. The GWAS dataset was evaluated using the ICSNPPathway, which showed seven candidate SNPs, five genes, and seven pathways that may contribute to the susceptibility of patients to ESCC.

Keywords

Esophageal squamous cell carcinoma; Genome-wide association study; Pathway analysis

1. Introduction

Esophageal cancer (EC) is the sixth leading cause of cancer-related mortality and the eighth most frequently diagnosed cancer worldwide. Approximately more than 450,000 people are afflicted with EC, which shows a rapidly increasing incidence rate (Zhu et al., 2014). Two major histological types of esophageal cancer (EC) have been found, namely, esophageal squamous cell carcinoma (ESCC) and esophageal adenocarcinoma. ESCC is the predominant histological type of EC world-wide, particularly in the so-called Asian belt

*Corresponding author at: Department of Radiation Oncology, The First Affiliated Hospital of Nanjing Medical University, 300 Guangzhou Road, Nanjing 210029, China. address: sunxinchun2012@163.com (X. Sun).

¹These authors contributed equally to this work.

Conflict of interest: We declare that we have no conflict of interest.

(Turkey, northeastern Iran, Kazakhstan, and northern and central China), where a significantly high incidence rate of ESCC has been noted accounting for approximately 90% of the total EC cases (Kamangar et al., 2006). Although ESCC is complex and heterogeneous and its etiology has not yet been determined, case-control and family studies have established a genetic component of the susceptibility to the disease (Kamangar et al., 2006). Single nucleotide polymorphisms (SNPs) are highly associated with the risk of ESCC within genes that encode proteins such as P53, PTEN, HIF-1 α , VEGF, ATM, and survivin (Guo et al., 2014; Xu et al., 2014a; Yang et al., 2014).

Genome-wide association studies (GWASs) offer a powerful approach in searching for genes that confer susceptibility to complex diseases (Wang et al., 2010). An increasing number of GWAS reports have led to the discovery and validation of disease-causing genes (Lee et al., 2012). Although large-scale GWASs have been conducted on complex diseases, including ESCC, several genetic components that contribute to ESCC variations remain ambiguous. One of the key challenges in GWAS data interpretation is the identification of causative SNPs and provision of evidence and hypothetical mechanisms responsible for the observed traits (Ge et al., 2014; Lee et al., 2014). Thus, new methods have been applied to study the existing GWAS data that may provide additional biological insights and highlight new candidate genes. The identify candidate causal SNPs and pathways (ICSNPPathway) analysis has been developed to identify candidate SNPs and their corresponding candidate pathways using GWAS data and by integrating linkage disequilibrium (LD) analysis, functional SNP annotation, and pathway-based analysis (PBA) (Zhang et al., 2011).

Accordingly, the ICSNPPathway analysis was used to search an ESCC GWAS dataset (Li et al., 2013) for candidate causal SNPs and candidate causal mechanisms of ESCC to generate SNP-to-gene-to-pathway hypotheses.

2. Materials and methods

2.1. Study population

A publicly available ESCC GWAS dataset from NCBI dbGap (study accession: phs000361.v1.p1) was explored. The dataset is based on a GWAS of ESCC at the NCI's Core Genotyping Facility with the Illumina 660W Quad chip in 2100 controls and 1898 ESCC cases (Li et al., 2013). The ESCC cases were from two studies conducted in north central China and included a single case-control and a cohort study. The dataset was filtered to remove individuals with $p < 0.001$ for Hardy-Weinberg violation and a call rate of $< 98\%$ to reduce the effect of genotyping errors. A total of 463886 SNPs passed the quality control filters.

2.2. Identification of candidate causal SNPs and pathways

ICSNPPathway analysis was performed in two stages. The first stage involved the pre-selection of candidate causal SNPs using LD analysis and functional SNP annotation based on the most significant SNPs. The second stage involved the annotation of biological mechanisms to the pre-selected candidate causal SNPs using the PBA algorithm improved gene set enrichment analysis (*i*-GSEA). A full list of ESCC GWAS SNP p values was introduced into the ICSNPPathway analysis. One concept applied in this process was LD

analysis, which searches for the most significant SNPs in LD within a GWAS dataset to identify more possible candidate causal SNPs based on an extended dataset that includes HapMap data. The other method involves the use of functional SNPs. ICSNPathway analysis pre-selects candidate causal SNPs based on functional SNPs, which are important for understanding the underlying genetics of human health. Functional SNPs are defined as SNPs that may alter protein or gene expression or the role of a protein in a pathway. They include deleterious and non-deleterious non-synonymous SNPs that cause the gain or loss of a stop codon and a frame shift as well as SNPs located in essential splice sites or in regulatory regions. The ICSNPathway server applies the *i*-GSEA PBA algorithm to the full list of the GWAS SNP p values to detect the pathways associated with individual traits. The process involves five steps. First, each SNP is mapped to its nearest gene according to the localization of the SNP and the gene in the Ensembl 61 database (<http://www.ensembl.org/biomart/martview>). The maximum $t = -\log(p \text{ value})$ values of the SNPs mapped to the genes are then assigned to represent those genes. All genes are subsequently ranked by decreasing representative t values. Second, for each pathway S , the enrichment score [ES , i.e., a Kolmogorov–Smirnov-like running sum statistics with weight (a)] is calculated. This procedure measures the tendency for genes of a pathway to be located at the top of the ranked gene list. Third, the ES is then converted to a significant-proportion-based ES ($SPES$) by multiplying with m_1/m_2 , where m_1 is the proportion of the significant genes for pathway S (defined as genes mapped with at least one SHLP in the top 5% of the most significant SNPs in the GWAS), and m_2 is the proportion of the significant genes among all genes in the GWAS. Fourth, SNP label permutation and normalization are performed to generate the distribution of $SPES$ and correct for gene (bias caused by different genes with different numbers of mapped SNPs) and pathway (bias caused by different pathways with different numbers of genes) variations. Last, based on the distribution of the $SPES$ values generated through the permutation, a nominal p value is calculated, and a false discovery rate (FDR) is computed for multiple-testing correction.

The term “the most significant SNP” refers to SNPs with p values below a certain threshold, which can be specified from the GWAS SNP p values. The ICSNPathway was used to analyze the significant pathways from the original GWAS when the p value threshold (<0.001) used in the study was selected. Two parameters were set for the analysis. The first parameter required analysis of 100 kb upstream and downstream of the gene, suggesting that only the p values of SNPs located within genes were used in the PBA algorithm. The second parameter was an FDR cutoff (0.05) for multiple-testing corrections. Controlling the FDR is preferred for large-scale testing. Defined as the expected proportion of false positives among all significant tests, the FDR allows researchers to identify a set of “candidate positives,” a large proportion of which is likely to be true positives. The FDR, a permutation-based approach for multiple comparisons, was used to identify the statistically significant genes. No specific criterion for selecting the number of genes was found. A minimum of 5 and maximum of 100 were used as cutoffs to avoid significantly narrow or broad functional categories. Pathways that contained more than 100 genes were discarded to avoid stochastic bias and inclusion of a general biological process. From the several options available for pathway annotation, four pathway databases were selected, namely, the Kyoto Encyclopedia of Genes and Genomes (KEGG), gene ontology (GO) biological process database, BioCarta

database and GO molecular function database. This selection ensured comprehensive coverage of pathways and high-quality information for well-defined pathways.

The SNP Annotation and Proxy (SNAP) search method has been developed to identify and annotate nearby SNPs in LD (proxies) using HapMap analysis (<http://www.broadinstitute.org/mpg/snap/>). In the present study, SNAP was used to find proxy SNPs, determine whether SNP proxies were present in genes, resolve whether associations from multiple SNPs represented similar associations, plot regional views of associations or LD structures, and retrieve annotations for SNPs.

3. Results

3.1. Candidate SNPs and pathways resulting from the ESCC GWAS

With the use of 463886 GWAS SNP p values as input and the most significant SNPs ($p < 0.001$), ICSNPPathway analysis identified seven candidate SNPs, five genes, and five pathways (Tables 1–3).

The top four candidate SNPs were rs1800450 [$-\log_{10}(p) = 2.218$], rs3769823 [$-\log_{10}(p) = 2.610$], rs3765524 [$-\log_{10}(p) = 7.026$], and rs2274223 [$-\log_{10}(p) = 6.924$]. Two of the five candidate SNPs, i.e., rs1800450 and rs3769823, were not in LD with any SNP. SNP rs4135113, which was not represented in the original GWAS metaanalysis, was in LD with rs4135054 ($r^2 = 1.0$). rs11187870 and rs647126, which were not represented in the original GWAS metaanalysis, were in LD with rs3781264 and rs590336 ($r^2 = 1.0$).

The biological mechanisms indicate that the candidate SNP may alter the role of its corresponding gene/protein involved in the pathway(s) associated with traits. The seven candidate SNPs included in seven candidate pathways represent five hypothetical biological mechanisms. For instance, the strongest hypothetical biological mechanism involved the role alteration of thymine DNA glycosylase (TDG) by rs4135113 in the pathways of non-synonymous coding (deleterious) in base excision repair (nominal $p < 0.001$; FDR = 0.002). The second strongest mechanism involved rs1800450 in the monosaccharide binding pathway (nominal $p < 0.001$; FDR = 0.002). The third mechanism involved rs3769823 of CASP8 in the D4-GDI signaling pathway (nominal $p < 0.001$; FDR = 0.002). The fourth mechanism involved rs11187870, rs3765524, and rs2274223 of the PLCE1 gene in the lipid bio-synthetic process pathway (nominal $p < 0.001$; FDR = 0.005). The fifth mechanism involved rs647126 of UCP3 in the respiratory gaseous exchange pathway (nominal $p < 0.001$; FDR = 0.006, Tables 1–3).

The genes that had a role in the pathways revealed by the pathway-based approach were investigated, and a distinct clustering of genes involved in the seven candidate causal pathways was found (Table 1). The most significant pathway was the positive regulation of a base excision repair, and the following genes were involved in this pathway: SMAD3, RHOA, NLRP12, CD27, and TNF ($p < 0.05$). The second significant pathway was the monosaccharide binding pathway, and the following genes were involved in this pathway: CASP3, CASP7, BIK, BCL2, DFFB, BAK1, CASP9, BID, CASP8, CASP6, BCL2L1, APAF1, BIRC3, and BIRC2 ($p < 0.05$). Genes without known immunological functions are

of considerable interest because they could elucidate the mechanisms for ESCC susceptibility. This analysis demonstrated that the pathways and genes found through PBA may contribute to ADHD susceptibility.

4. Discussion

Multiple related genes in a pathway may work together to confer susceptibility to cancer, but not all of these genes could achieve genome-wide significance in any single GWAS.

Therefore, pathway-based analysis is required to identify new loci associated with susceptibility to cancer risk (Lee et al., 2014). Various cellular pathways and complex molecular networks may have key roles in the ESCC development. If a specific pathway is relevant to cancer susceptibility, the associated signals are expected to be over-represented for SNPs in that pathway (Wang et al., 2014). Given that the power of GWAS to detect single SNP associations is limited, a pathway-based approach was used to account for the biological association between genes and provide insights into how multiple genes might contribute to the pathogenesis of ESCC.

In the present study, ICSNPathway analysis was conducted to identify the candidate SNPs, genes, and pathways that suggested the hypothetical biological mechanisms that affect the susceptibility to ESCC. The most significant SNP-to-gene-to-effect hypothesis was that rs4135113 alters the role of TDG in base excision repair pathway. TDG is a glycosylase to correct G/T or G/U mismatches in BER pathway. It has been shown that TDG might also be involved in gene transcription, such as nuclear receptor signaling. A mutation of TDG will contribute to colorectal cancer based on various analyses. As a coactivator, it could promote beta-catenin/TCF transactivation and functionally cooperates with CBP in Wnt signaling, knocking down TDG could inhibit the proliferation of colon cells (Jia et al., 2014; Xu et al., 2014b), which promoted that TDG may contribute to the risk of digestive cancer. In a previous GWAS study, TDG SNP (rs4135054) was associated with the risk of ESCC (Li et al., 2013). However, no close linkage between these previously reported SNPs and rs4135054 (TDG) through LD search was found in that study (Li et al., 2013). In the present study, a candidate causal SNP rs4135113 was found using pathway analysis, suggesting that TDG may have an important role in ESCC risk. The second strongest mechanism involved the modulation of monosaccharide binding that selectively interacts with any monosaccharide by rs1800450, which may alter the expression of Mannan-Binding Lectin 2 (MBL2). Mannan-binding lectin (MBL), which encodes the soluble mannose-binding lectin or mannose-binding protein found in serum, is a member of the collectin family and plays an important role in the innate immune system (Swierzko et al., 2014). It can act directly as an opsonin or activating MBL-associated serine proteases (MASPs), thus initiating the antibody-independent pathway of the complement system. The *MBL2* gene (*MBL1* is a pseudo gene) is located on chromosome 10q11.2-q21 and comprises four exons (te Poele et al., 2012). In previous study, rs1800450 was associated with a higher risk of glioma (Michaud et al., 2013), here we first report that rs1800450 may contribute to the risk of ESCC.

Thus, five candidate genes (*TDG*, *MBL2*, *CASP8*, *PLCE1*, and *UCP3*) that may contribute to increased risk of ESCC have been identified. Previous GWAS and association studies have

identified that most of the candidate genes found in the present analysis affect susceptibility to ESCC. Although the hypothetical mechanisms proposed in the present study might contribute to the susceptibility to ESCC, the PBA test may also be prone to false positive results, which are often associated with single-marker-based association tests. Therefore, these results should be viewed as preliminary to generate new hypotheses that should then be appropriately verified through independent studies. Thus, additional studies are needed to confirm the association among the candidate SNPs, genes, and pathways identified in the present study and their roles in ESCC. Nevertheless, pathway-based approaches have complementary roles in the identification of genes that confer disease susceptibility. Therefore, the results obtained in the present study may lead to the formulation of novel hypotheses for future investigation.

In summary, an ESCC GWAS dataset was examined to identify genetic associations with ESCC at both SNP and pathway levels. The application of ICSNPathway analysis on the ESCC GWAS dataset led to the identification of seven candidate SNPs and five genes (including *TDG*, *MBL2*, *CASP8*, *PLCE1*, and *UCP3*), seven pathways, and seven biological mechanisms that may contribute to ESCC cancer susceptibility. However, further studies are needed to confirm and explore the genetic variations of the molecular pathways that may be associated with ESCC.

Acknowledgments

This study was supported by the Natural Science Foundation of China (No. 81272504), the Innovation Team (No. LJ201123-EH11), Jiangsu Provincial Science and Technology Projects BK2011854 (DA11), the Six Major Talent Peak Project of Jiangsu Province (2013-WSN-040), a project funded by the priority academic program development of Jiangsu Higher Education Institution (JX10231801), grants from Key Academic Discipline of Jiangsu Province “Medical Aspects of Specific Environments”, and “333” Project of Jiangsu Province BRA2012210 (RS12).

References

- Ge YZ, Xu Z, Xu LW, Yu P, Zhao Y, Xin H, Wu R, Tan SJ, Song Q, Wu JP, Li WC, Zhu JG, Jia RP. Pathway analysis of genome-wide association study on serum prostate-specific antigen levels. *Gene*. 2014; 551:86–91. [PubMed: 25168891]
- Guo Q, Dai SB, Shen F, Yu D, Shen ST, Zhang Q, Huang JX, Wu ZD. VEGF +405G/C (rs2010963) polymorphisms and digestive system cancer risk: a meta-analysis. *Tumour Biol*. 2014; 35:4977–4982. [PubMed: 24474253]
- Jia Y, Nie F, Du A, Chen Z, Qin Y, Huang T, Song X, Li L. Thymine DNA glycosylase promotes transactivation of beta-catenin/TCFs by cooperating with CBP. *J Mol Cell Biol*. 2014; 6:231–239. [PubMed: 24748645]
- Kamangar F, Dores GM, Anderson WF. Patterns of cancer incidence, mortality, and prevalence across five continents: defining priorities to reduce cancer disparities in different geographic regions of the world. *J Clin Oncol*. 2006; 24:2137–2150. [PubMed: 16682732]
- Lee YH, Bae SC, Choi SJ, Ji JD, Song GG. Genome-wide pathway analysis of genome-wide association studies on systemic lupus erythematosus and rheumatoid arthritis. *Mol Biol Rep*. 2012; 39:10627–10635. [PubMed: 23053960]
- Lee YH, Kim JH, Song GG. Genome-wide pathway analysis in neuroblastoma. *Tumour Biol*. 2014; 35:3471–3485. [PubMed: 24293394]
- Li WQ, Hu N, Hyland PL, Gao Y, Wang ZM, Yu K, Su H, Wang CY, Wang LM, Chanock SJ, Burdett L, Ding T, Qiao YL, Fan JH, Wang Y, Xu Y, Shi JX, Gu F, Wheeler W, Xiong XQ, Giffen C, Tucker MA, Dawsey SM, Freedman ND, Abnet CC, Goldstein AM, Taylor PR. Genetic variants in

DNA repair pathway genes and risk of esophageal squamous cell carcinoma and gastric adenocarcinoma in a Chinese population. *Carcinogenesis*. 2013; 34:1536–1542. [PubMed: 23504502]

Michaud DS, Siddiq A, Cox DG, Backes DM, Calboli FC, Sughrue ME, Gaziano JM, Ma J, Stampfer M, Tworoger SS, Hunter DJ, Camargo CA Jr, Parsa AT. Mannose-binding lectin 2 gene and risk of adult glioma. *PLoS One*. 2013; 8:e61117. [PubMed: 23637788]

Swiercko AS, Szala A, Sawicki S, Szemraj J, Sniadecki M, Sokolowska A, Kaluzynski A, Wydra D, Cedzynski M. Mannose-Binding Lectin (MBL) and MBL-associated serine protease-2 (MASP-2) in women with malignant and benign ovarian tumours. *Cancer Immunol Immunother*. 2014; 63(11): 1129–1140. [PubMed: 25038892]

te Poele EM, Siedlinski M, Anne de Pagter PJ, Bierings MB, Scherpen FJ, Meeuwsende Boer TG, Koppelman GH, Postma DS, Kamps WA, Boezen HM, de Bont ES. MBL2 and fever during neutropenia in children with acute lymphoblastic leukaemia. *Br J Haematol*. 2012; 157:132–135. [PubMed: 22050689]

Wang K, Li M, Hakonarson H. Analysing biological pathways in genome-wide association studies. *Nat Rev Genet*. 2010; 11:843–854. [PubMed: 21085203]

Wang AH, Liu Y, Wang B, He YX, Fang YX, Yan YP. Epidemiological studies of esophageal cancer in the era of genome-wide association studies. *World J Gastrointest Pathophysiol*. 2014; 5:335–343. [PubMed: 25133033]

Xu L, Zhou X, Xu L, Yin R. Survivin rs9904341 (G>C) polymorphism contributes to cancer risk: an updated meta-analysis of 26 studies. *Tumour Biol*. 2014a; 35:1661–1669. [PubMed: 24096544]

Xu X, Yu T, Shi J, Chen X, Zhang W, Lin T, Liu Z, Wang Y, Zeng Z, Wang C, Li M, Liu C. Thymine DNA glycosylase is a positive regulator of Wnt signaling in colorectal cancer. *J Biol Chem*. 2014b; 289:8881–8890. [PubMed: 24532795]

Yang X, Zhang C, Zhu HC, Qin Q, Zhao LJ, Liu J, Xu LP, Zhang Q, Cai J, Ma JX, Cheng HY, Sun XC. HIF-1alpha P582S and A588T polymorphisms and digestive system cancer risk—a meta-analysis. *Tumour Biol*. 2014; 35:2825–2830. [PubMed: 24293391]

Zhang K, Chang S, Cui S, Guo L, Zhang L, Wang J. ICSNPathway: identify candidate causal SNPs and pathways from genome-wide association study by one analytical framework. *Nucleic Acids Res*. 2011; 39:W437–W443. [PubMed: 21622953]

Zhu HC, Yang X, Xu LP, Zhao LJ, Tao GZ, Zhang C, Qin Q, Cai J, Ma JX, Mao WD, Zhang XZ, Cheng HY, Sun XC. Meat consumption is associated with esophageal cancer risk in a meat- and cancer-histological-type dependent manner. *Dig Dis Sci*. 2014; 59:664–673. [PubMed: 24395380]

Abbreviations

ESCC	esophageal squamous cell carcinoma
EC	esophageal cancer
SNP	single nucleotide polymorphism
GWAS	genome-wide association study
ICSNPathway	the identify candidate causal SNPs and pathways
LD	linkage disequilibrium
PBA	pathway-based analysis
NCBI	National Center for Biotechnology Information
HWE	Hardy–Weinberg equilibrium
QC	quality control
KEGG	the Kyoto Encyclopedia of Genes and Genomes

GO	gene ontology
SPES	significant proportion based enrichment score
FDR	false discovery rate
SD	standard deviation

Table 1

Candidate causal single nucleotide polymorphisms identified from ICSNPPathway analysis.

Candidate causal SNP	Functional class	Gene	Candidate causal pathway ^a	$-\log_{10}(p)$ ^b	In LD with	r^2	D'	$-\log_{10}(p)$ ^c
rs4135113	Non_synonymous_coding(deleterious)	<i>TDG</i>	1	-	rs4135054	1	1	3.065
rs1800450	Non_synonymous_coding(deleterious)	<i>MBL2</i>	2	2.218	rs2120132	1	1	3.67
rs3769823	Non_synonymous_coding	<i>CASP8</i>	3,7	2.61	rs13016963	0.943	1	5.214
rs11187870	Regulatory_region	<i>PLCE1</i>	4,5	-	rs3781264	0.857	1	6.939
rs3765524	Non_synonymous_coding	<i>PLCE1</i>	4,5	7.026	rs3765524	-	-	7.026
rs2274223	Non_synonymous_coding	<i>PLCE1</i>	4,5	6.924	rs2274223	-	-	6.924
rs647126	Regulatory_region	<i>UCP3</i>	6	-	rs590336	0.909	1	3.827

SNP single-nucleotide polymorphism, LD linkage disequilibrium

“.” denotes that this SNP was not represented in the original GWAS.

^aNumbers indicate the indices of pathways ranked by significance (false discovery rate).

^b $-\log_{10}(p)$ values of candidate causal SNPs in the original genome-wide association studies (GWASs).

^c $-\log_{10}(p)$ values of SNPs in LD with candidate causal SNPs in the original GWAS.

Table 2

Candidate causal pathways for esophageal squamous cell carcinoma.

Index	Candidate causal pathway	Description	Nominal p	FDR
1	BASE EXCISION REPAIR	GO:0006284. In base excision repair, an altered base is removed by a DNA glycosylase enzyme, followed by excision of the resulting sugar phosphate. The small gap left in the DNA helix is filled in by the sequential action of DNA polymerase and DNA ligase.	<0.001	0.002
2	MONOSACCHARIDE BINDING	GO:0048029. Interacting selectively with any monosaccharide. Monosaccharides are the simplest carbohydrates; they are polyhydric alcohols containing either an aldehyde or a keto group and between three to ten or more carbon atoms. They form the constitutional repeating units of oligo- and polysaccharides.	<0.001	0.002
3	d4gdiPathway	D4-GDI Signaling Pathway	<0.001	0.002
4	LIPID BIOSYNTHETIC PROCESS	GO:0008610. The chemical reactions and pathways resulting in the formation of lipids, compounds soluble in an organic solvent but not, or sparingly, in an aqueous solvent.	<0.001	0.005
5	REGULATION OF CELL GROWTH	GO:0001558. Any process that modulates the frequency, rate or extent of cell growth.	<0.001	0.006
6	RESPIRATORY GASEOUS EXCHANGE	GO:0007585. The process of gaseous exchange between an organism and its environment. In plants, microorganisms, and many small animals, air or water makes direct contact with the organism's cells or tissue fluids, and the processes of diffusion supply the organism with dioxygen (O ₂) and remove carbon dioxide (CO ₂). In larger animals the efficiency of gaseous exchange is improved by specialized respiratory organs, such as lungs and gills, which are ventilated by breathing mechanisms.	<0.001	0.006
7	caspasePathway	Caspase Cascade in Apoptosis	<0.001	0.006

PN, nominal P value; FDR, false discovery rate; GO, gene ontology.

Table 3

Function and association study of genes identified by GWAS pathway analysis.

TDG

The protein encoded by this gene belongs to the TDG/mug DNA glycosylase family. Thymine-DNA glycosylase (TDG) removes thymine moieties from G/T mismatches by hydrolyzing the carbon–nitrogen bond between the sugar-phosphate backbone of DNA and the mispaired thymine. With lower activity, this enzyme also removes thymine from C/T and T/T mispairings. TDG can also remove uracil and 5-bromouracil from mispairings with guanine. This enzyme plays a central role in cellular defense against genetic mutation caused by the spontaneous deamination of 5-methylcytosine and cytosine. This gene may have a pseudogene in the p arm of chromosome 12.

MBL2

This gene encodes the soluble mannose-binding lectin or mannose-binding protein found in serum. The protein encoded belongs to the collectin family and is an important element in the innate immune system. The protein recognizes mannose and N-acetylglucosamine on many microorganisms, and is capable of activating the classical complement pathway. Deficiencies of this gene have been associated with susceptibility to autoimmune and infectious diseases.

CASP8

This gene encodes a member of the cysteine-aspartic acid protease (caspase) family. Sequential activation of caspases plays a central role in the execution-phase of cell apoptosis. Caspases exist as inactive proenzymes composed of a prodomain, a large protease subunit, and a small protease subunit. Activation of caspases requires proteolytic processing at conserved internal aspartic residues to generate a heterodimeric enzyme consisting of the large and small subunits. This protein is involved in the programmed cell death induced by Fas and various apoptotic stimuli. The N-terminal FADD-like death effector domain of this protein suggests that it may interact with Fas-interacting protein FADD. This protein was detected in the insoluble fraction of the affected brain region from Huntington disease patients but not in those from normal controls, which implicated the role in neurodegenerative diseases. Many alternatively spliced transcript variants encoding different isoforms have been described, although not all variants have had their full-length sequences determined.

PLCE1

This gene encodes a phospholipase enzyme that catalyzes the hydrolysis of phosphatidylinositol-4,5-bisphosphate to generate two second messengers: inositol 1,4,5-triphosphate (IP3) and diacylglycerol (DAG). These second messengers subsequently regulate various processes affecting cell growth, differentiation, and gene expression. This enzyme is regulated by small monomeric GTPases of the Ras and Rho families and by heterotrimeric G proteins. In addition to its phospholipase C catalytic activity, this enzyme has an N-terminal domain with guanine nucleotide exchange (GEF) activity. Mutations in this gene cause early-onset nephrotic syndrome; characterized by proteinuria, edema, and diffuse mesangial sclerosis or focal and segmental glomerulosclerosis. Alternative splicing results in multiple transcript variants encoding distinct isoforms.

UCP3

Mitochondrial uncoupling proteins (UCP) are members of the larger family of mitochondrial anion carrier proteins (MACP). UCPs separate oxidative phosphorylation from ATP synthesis with energy dissipated as heat, also referred to as the mitochondrial proton leak. UCPs facilitate the transfer of anions from the inner to the outer mitochondrial membrane and the return transfer of protons from the outer to the inner mitochondrial membrane. They also reduce the mitochondrial membrane potential in mammalian cells. The different UCPs have tissue-specific expression; this gene is primarily expressed in skeletal muscle. This gene's protein product is postulated to protect mitochondria against lipid-induced oxidative stress. Expression levels of this gene increase when fatty acid supplies to mitochondria exceed their oxidation capacity and the protein enables the export of fatty acids from mitochondria. UCPs contain the three solcar protein domains typically found in MACPs. Two splice variants have been found for this gene.